Informatics Database System at the NIST Combinatorial Methods Center

Cher H. Davis

Wenhua Zhang Michael J. Fasolka

Alamgir Karim Eric J. Amis

Polymers Division

Materials Science and Engineering Laboratory

National Institute of Standards and Technology

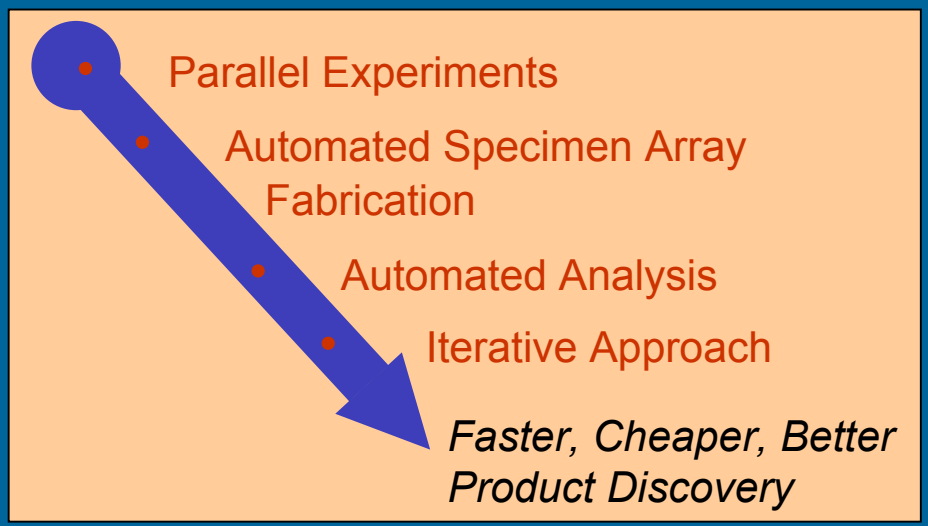
NIST
National Institute of Standards and Technology

Combinatorial and High-Throughput Techniques

Introduction

- Revolutionized the pharmaceutical industry
- Accelerating materials research

70% of the worlds 30 largest chemical companies have substantial investment in combinatorial programs"
- Peter Cohen, CTO, Symyx Corp. June, 2002




TAPPING INTO NIST'S COMBI EXPERTISE

C&E News article
Nov. 11, 2002
Vol. 80, #45
pp. 58-60

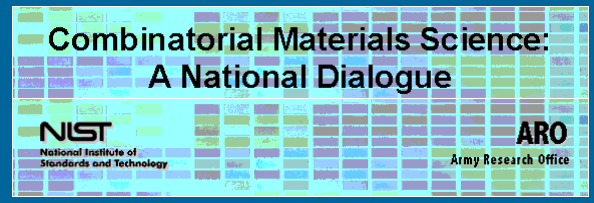
NIST National Institute of Standards and Technology

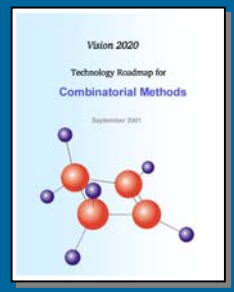
NIST Combinatorial Methods Center

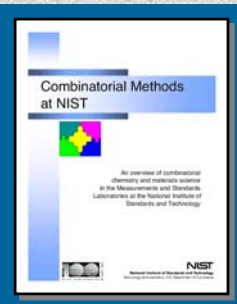
2000

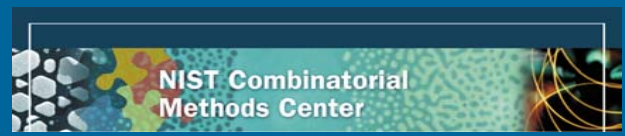
- Combinatorial Methods Road mapping Workshop, NIST, May, 2000
  - Vision 2020 "Technology Roadmap for Combinatorial Methods", September 2001
- Critical Issues in Designing and Characterizing Polymers for Medical Applications, NIST/New Jersey Center for Biomaterials, Rutgers, NJ June, 2000
- NIST Combi-Methods Center, July, 2001
- Combinatorial Methods Working Group (CMWG), NIST, Mar, 2002
  - Combinatorial Methods at NIST, NISTIR 6730
- COMBI 2002 Knowledge Foundation, NCMC kick-off, Jan, 2002
  - NCMC-1 "Library Design and Calibration", NIST, April, 2002
  - NCMC-2 "Combinatorial Adhesion and Mechanical Properties", NIST, Oct, 2002
  - NCMC-3 "Combinatorial Informatics", NIST, May, 2003
- Gordon Conference "Combinatorial and High Throughput Materials Science", Meriden, NH Jun, 2002
- COMBI 2003 Knowledge Foundation "High Throughput Materials Informatics", Feb, 2003

2003









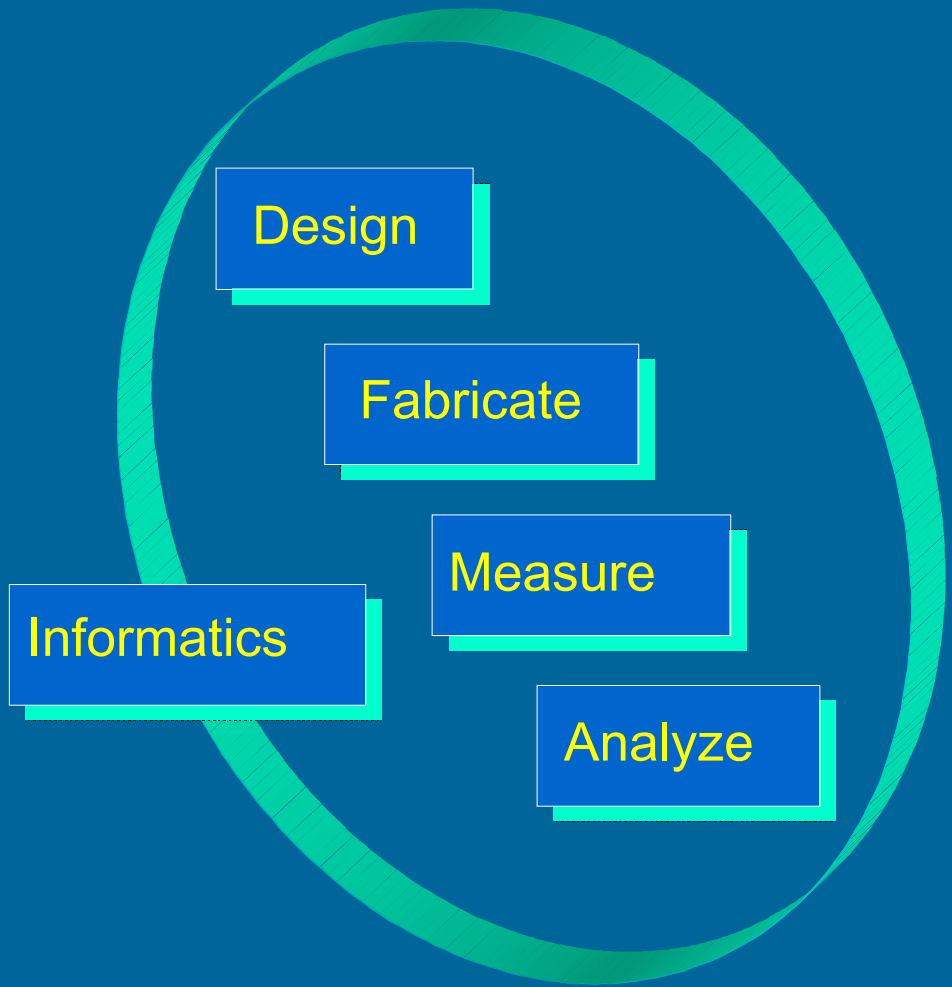
Technical Directions

- NIST Combinatorial Methods Center (NCMC)

- Films and Coatings
  - Topography and Wetting
  - Optical properties
  - Dielectric properties
- Lubricants, Adhesives
  - Tribology, Adhesion
  - Anisotropic properties
- Polymer Blends
  - Thermodynamics and kinetics
  - Phase separation morphology
  - Composition, temperature, processing
  - Additives (fillers, CO_2, process aids)
- Block Copolymers
  - Surface ordering
  - Surface energy
  - Temperature
- Biofunctional Materials
  - Biocompatibility
  - Topography
  - Cell adhesion
- Polymer Crystallization
  - Surface effects
  - Nucleating agents

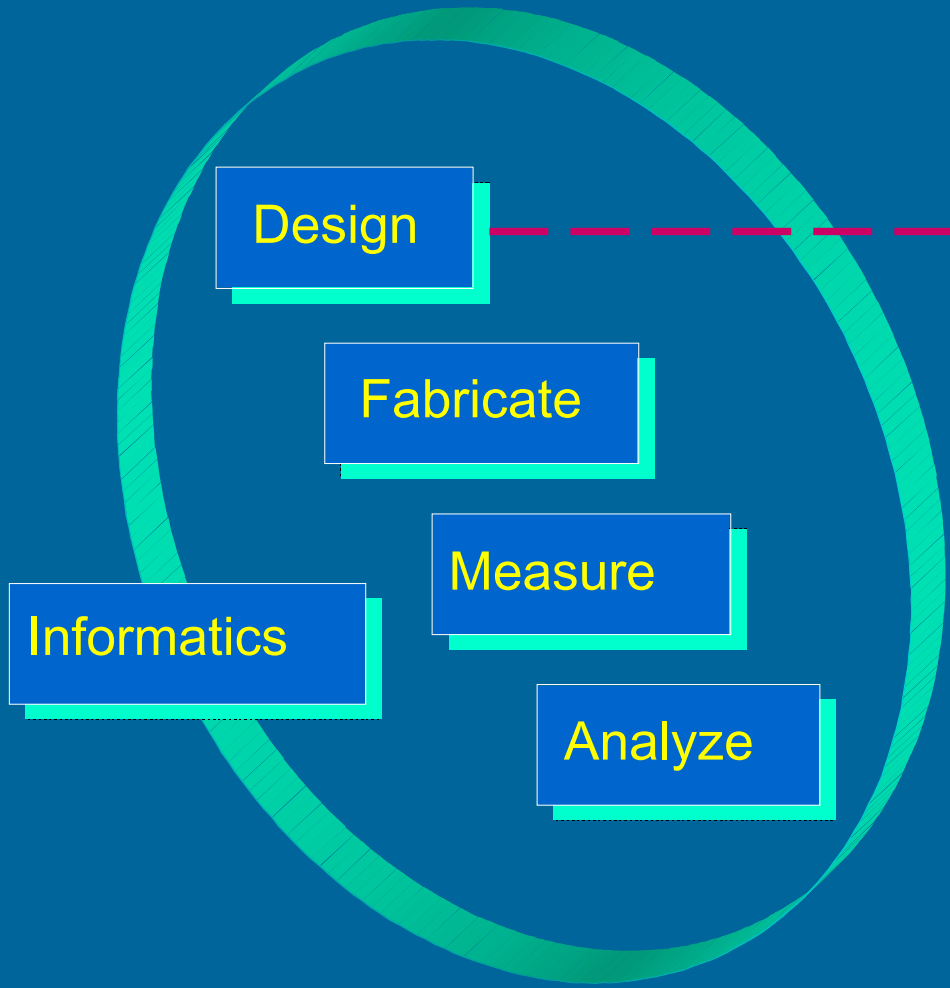
NIST
National Institute of Standards and Technology

The Combinatorial Cycle



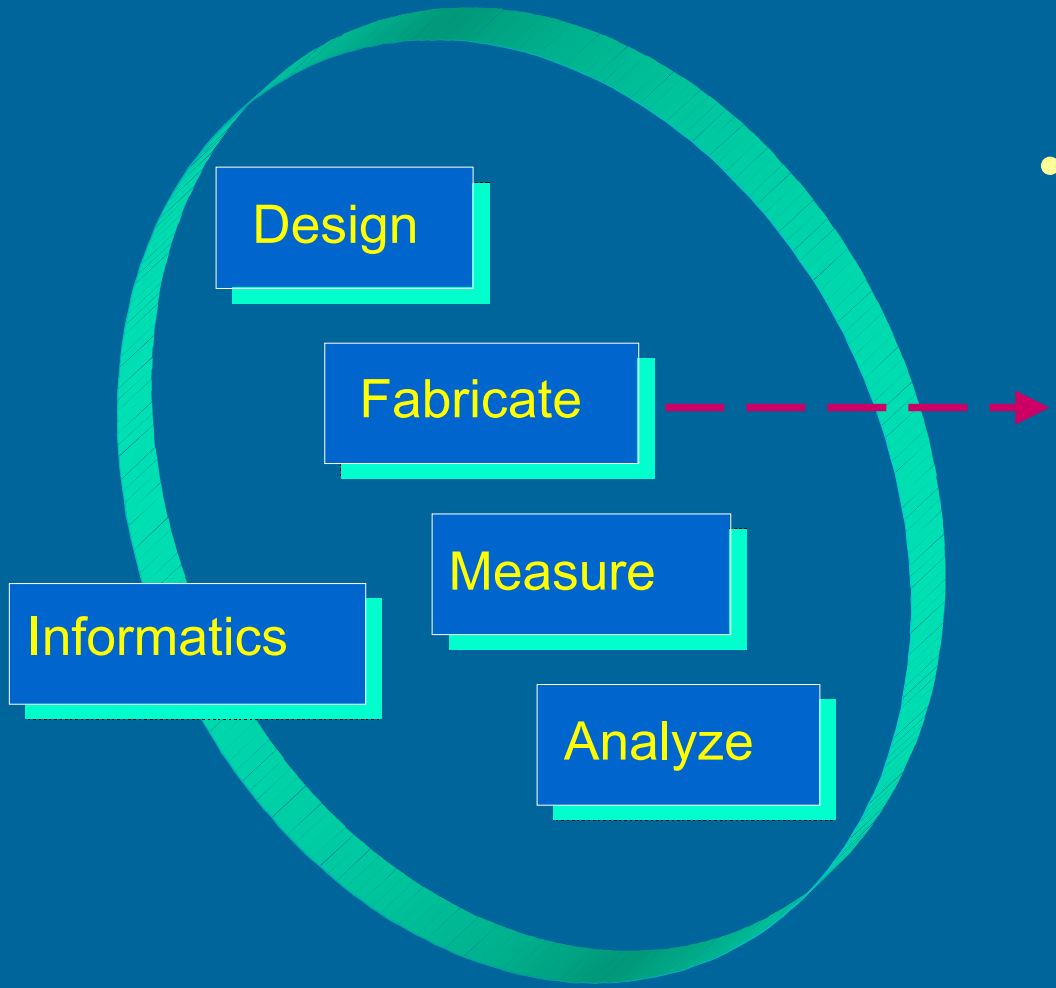
- Steps in Combinatorial research

Important Parameters



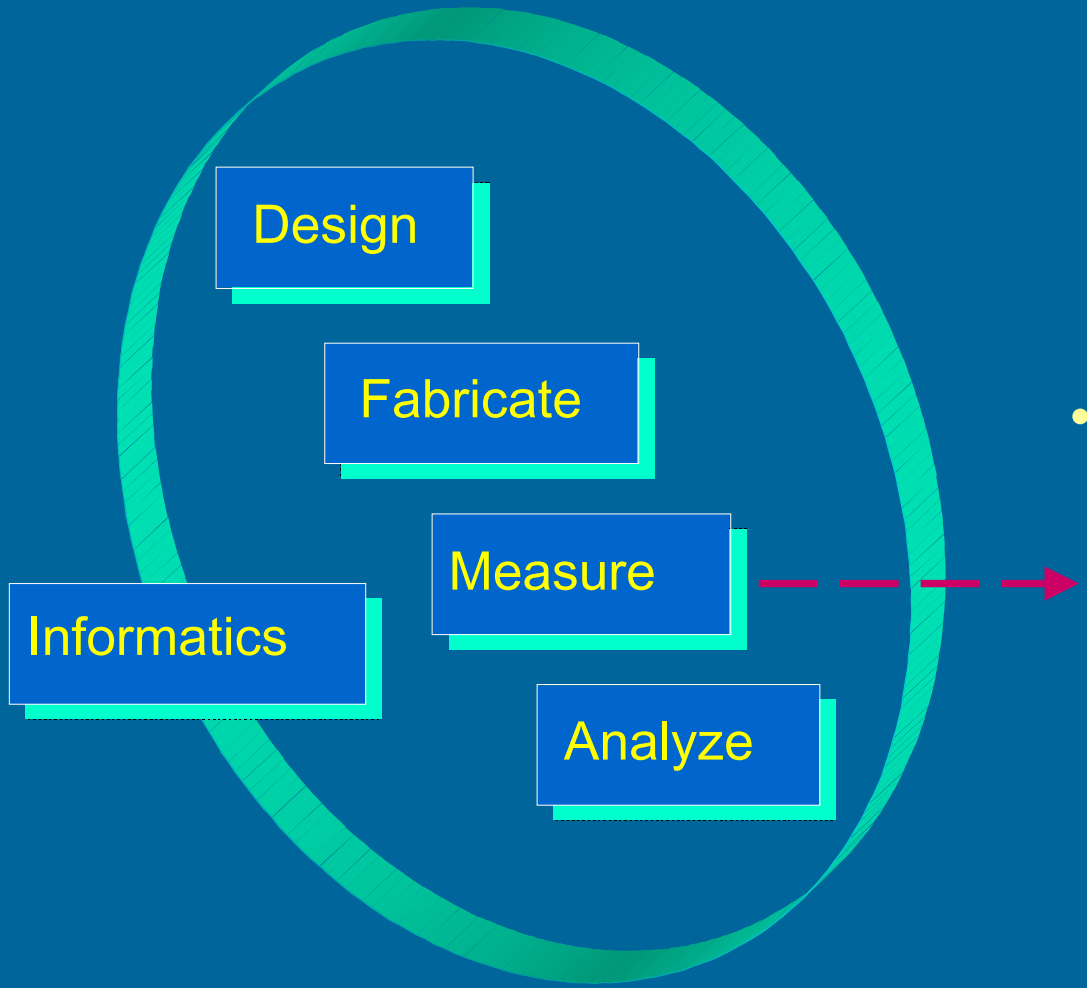
- Define sampling of parameter space
  - Molecular structure
  - Composition
  - Morphology
  - Processing

Library Fabrication



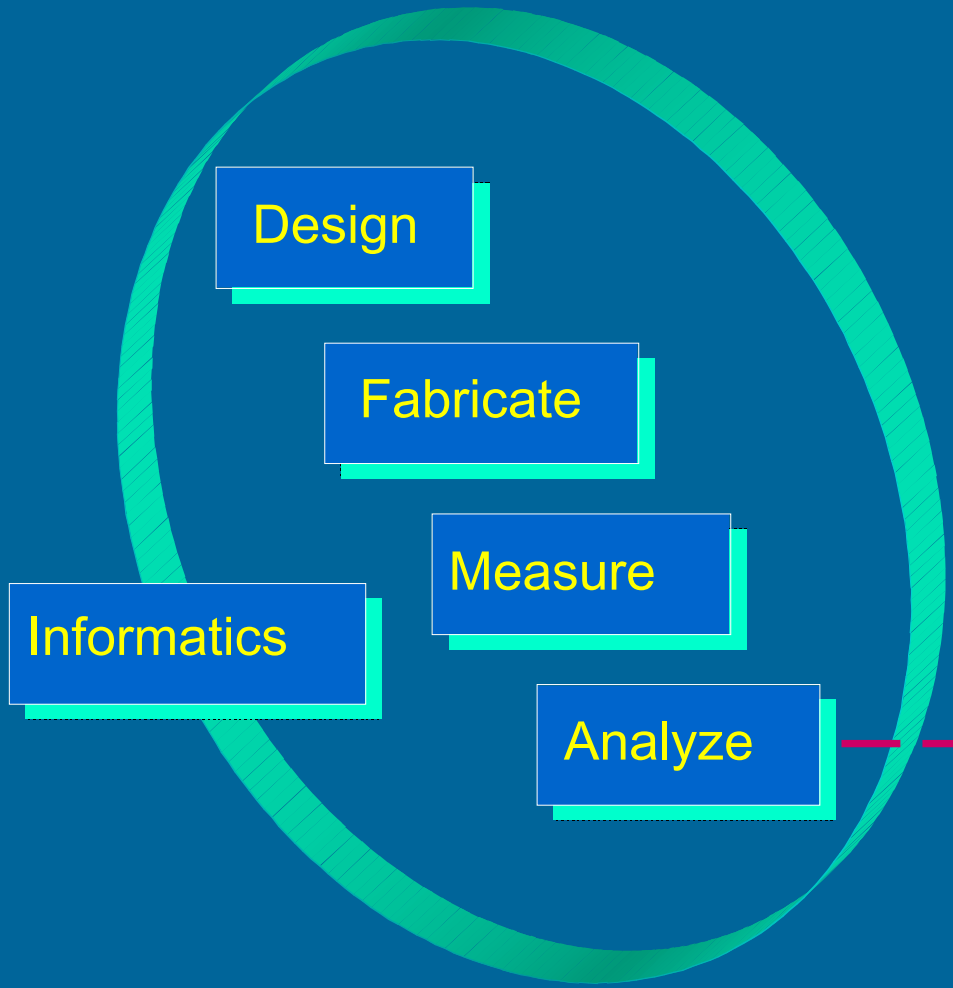
- Synthesis and creation of samples
  - Continuous gradients
  - Split and pool
  - Small/discrete samples
  - Patterning

High-Throughput Measurements



- Characterize materials, Discriminate performance
  - Screening hits
  - Comparative performance
  - Quantitative data

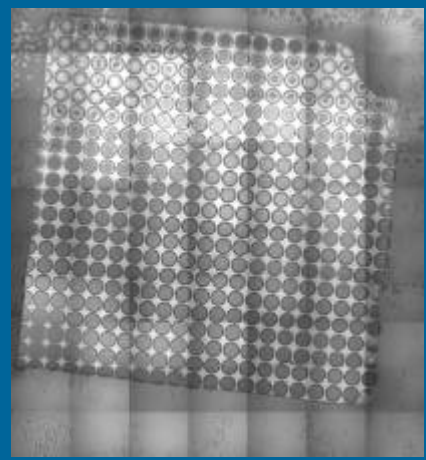
Automated Data Analysis

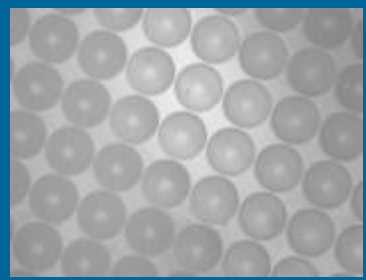


- From massive data sets to knowledge
  - Data management
  - Visualization
  - Data mining
  - Development of predictive models

Specific Example – Combinatorial Adhesion

- Library Design: Miniaturized multi-JKR (mm => μm)
- Library Fabrication: Micro-Lens array via soft-lithography
- High-Throughput Measurements: Rapid image acquisition with coordinated stage motion (Labview & Image-Pro)
- High-Throughput Analysis:
  - Extract contact radius (a) as a function of displacement (d)
  - Calculate the strain energy for each array displacement
  - Plot strain energy against crack velocity
  - Perform two parameter fit of experimental data to determine material properties

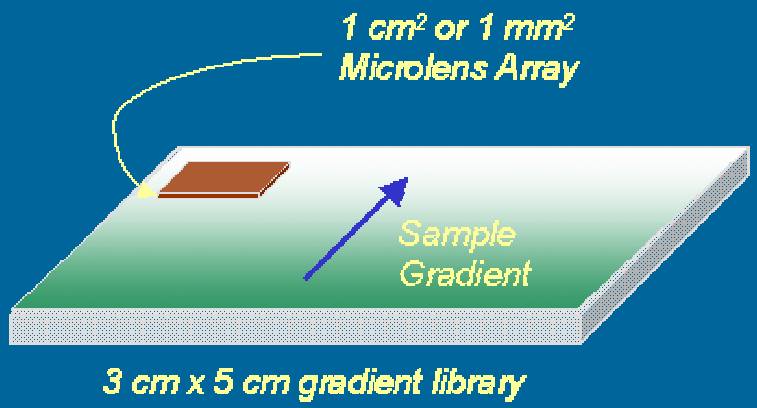




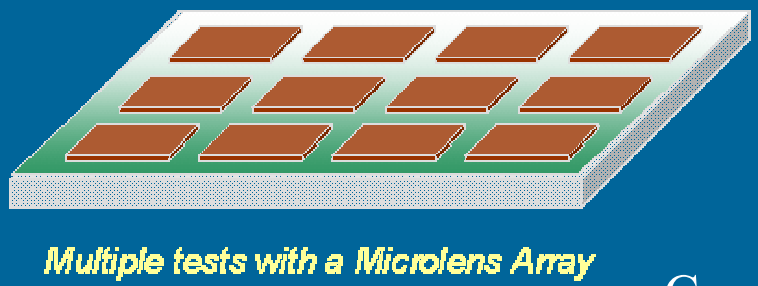
A. Crosby, C. Stafford, A. Forster

NIST National Institute of Standards and Technology

Need for Informatics



Single Microlens Test



Multiple Microlens Test

- Data generation
  - ~200 pictures
  - ~269 megabytes
  - Analysis: Fit displacement and area data for ~80 lenses
- Workload increases by a factor of n = 12
  - ~2400 pictures
  - ~3.30 gigabytes
  - Analysis: Fit displacement and area data for ~1000 lenses

Consider n = 100 => Info-flooding !!

Issues

Does data need to be stored locally or centrally?

Can informatics allow for real time data analyses?

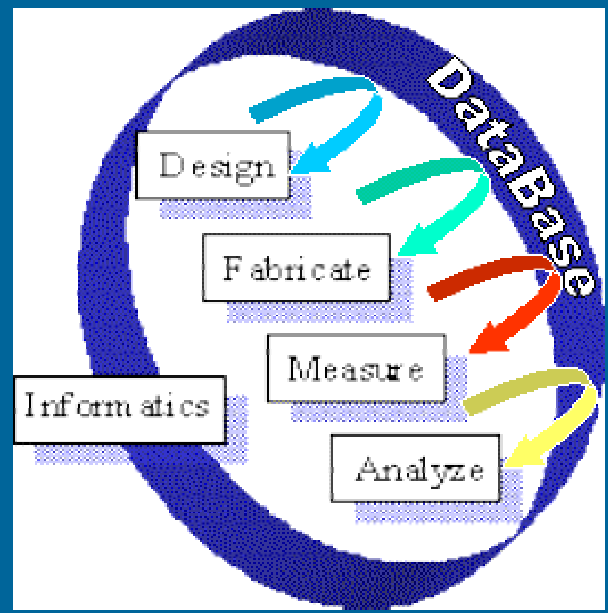
How about detecting flawed libraries early ?

NIST National Institute of Standards and Technology

Informatics—Closing the Loop

Informatics & feedback for:

- Data management, integration
- Data visualization
- Automated data entry, retrieval
- Protocols for standardized data formats
- Link to and refine other combi processes formats
- Library of scientific combi data



Informatics is needed at all stages of the combi cycle……

Informatics

Required at all Stages of the Combi-Cycle

- **Design of Experiments (DOE)**
  - Statistical models, Empirical data, Literature values, Prior data…
- **Fabrication of Libraries**
  - Instrument automation, Link to library database, Standardized experimental procedure…
- **High-throughput measurements**
  - Data management: automated data entry, retrieval, integration, consistent data format

    Data visualization …
- **High-throughput analysis**
  - Data processing, theoretical models, simulations, library of scientific combi-data…
- **Feedback**
  - Refine next set of HTE intelligently!

NCMC Informatics Project

- Consolidation/Integration of NCMC automation and analysis software
- Development of model combi research database system:

  Designed for transparency

  Easily understood structure

  Built on open source code

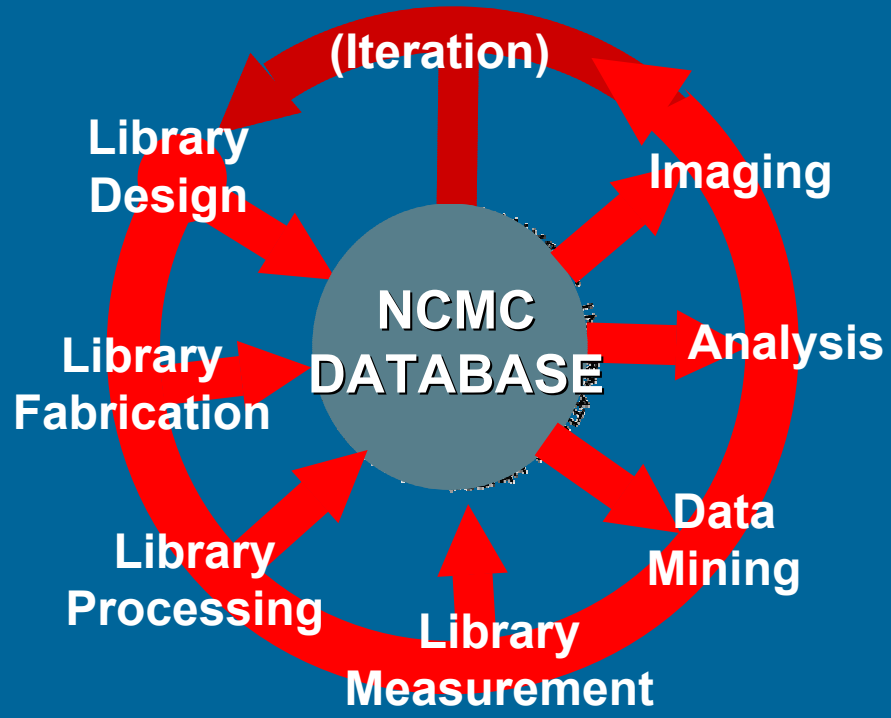
  Free to use and modify

  PostgreSQL database

  Python programming language

  Web Access

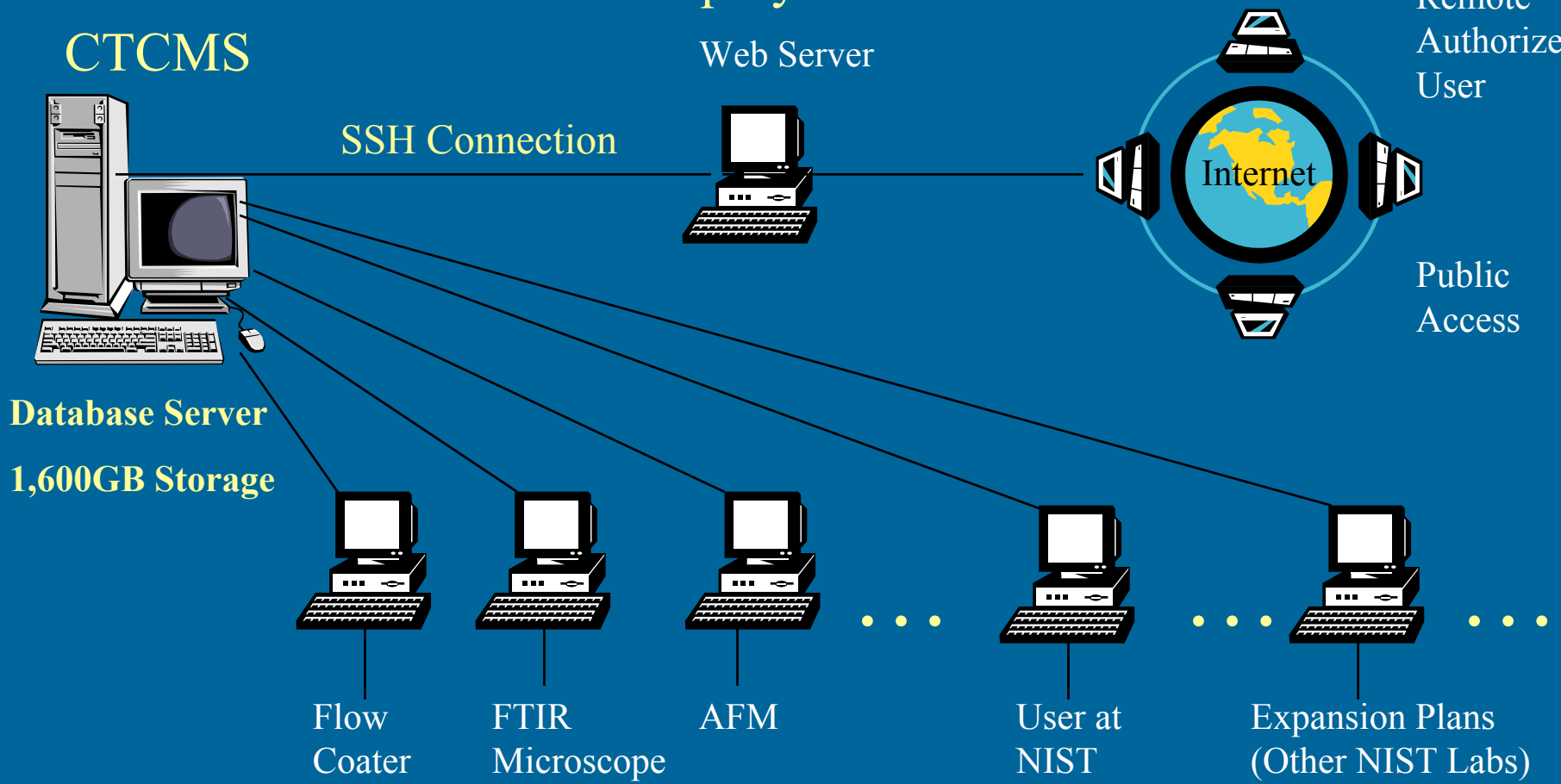
  Eventual public data mining



W. Zhang, M. Fasolka

NCMC Informatics Infrastructure Development

Initial Phase: Infrastructure Deployment



NCMC Instruments and Analysis Software

This phase is near completion

National Institute of Standards and Technology

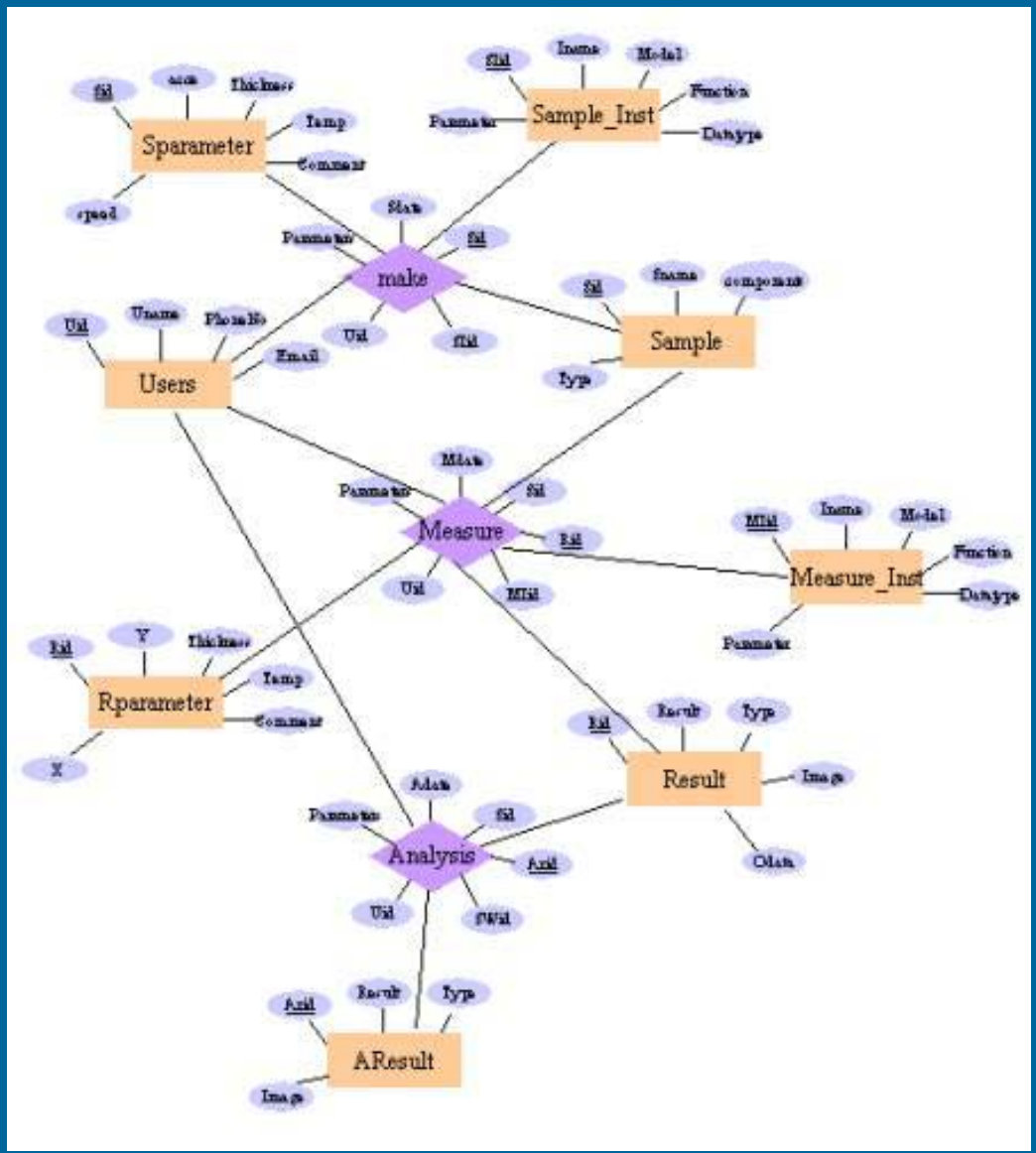
Database Management System PostgreSQL

- Sophisticated Object-Relational DBMS
- Supporting almost all SQL (Structured Query Language), including sub selects, transactions, and user-defined types and functions
- One of the most advanced open-source database available anywhere
- Commercial support available

For More Information:

www.postgresql.org

Basic Structure of Database

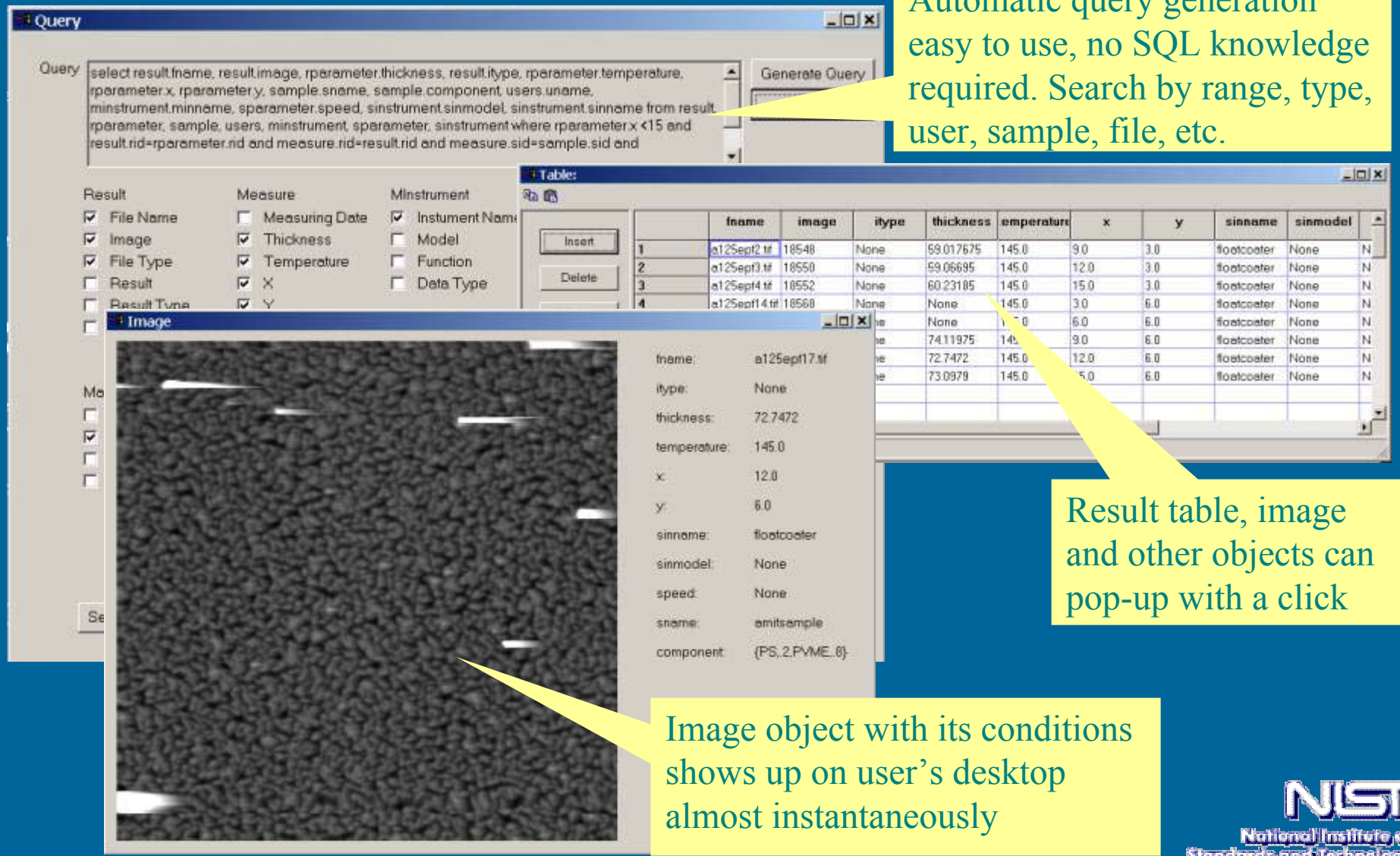


- Straight forward structure in consideration of automation & expansion
- Must reflect the frame-work of the Combi lab
- Establish basic protocol of Combi experiment set (All stages of Combi processes)
- Large objects, such as image, video, result chart, original data file storage

Software Technology

- Programming Language
  - Python (Open source, Object oriented)
- Secured Shell Connection with the Database server
- ZOPE selected as Web server
  - Base on Python, Open source, Easy to maintain, Database connectivity
- Integration with current data analysis software (mostly IDL)
- Consideration of future interface with other software

User Interface for Data Mining



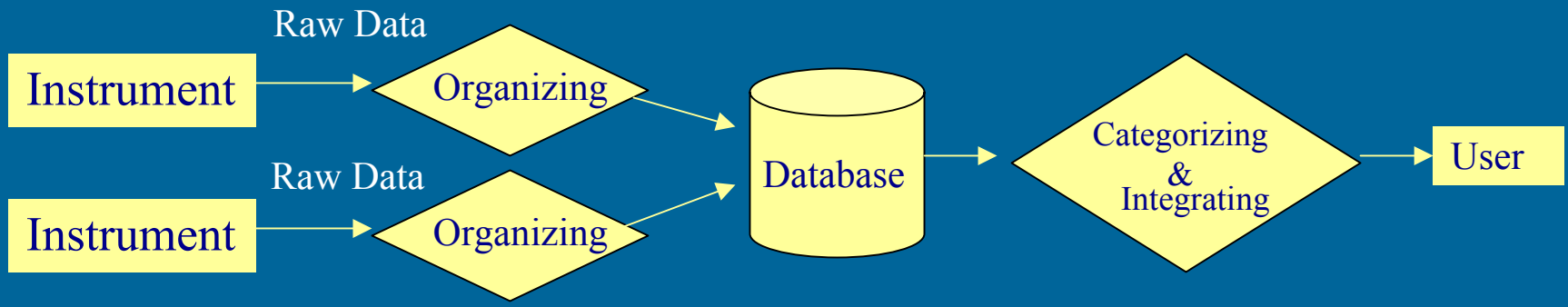
Automatic query generation easy to use, no SQL knowledge required. Search by range, type, user, sample, file, etc.

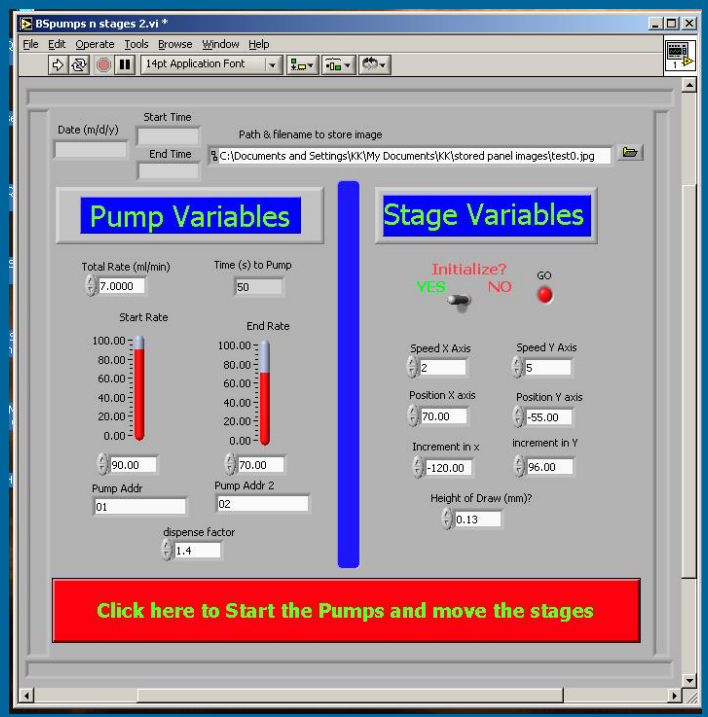
Result table, image and other objects can pop-up with a click

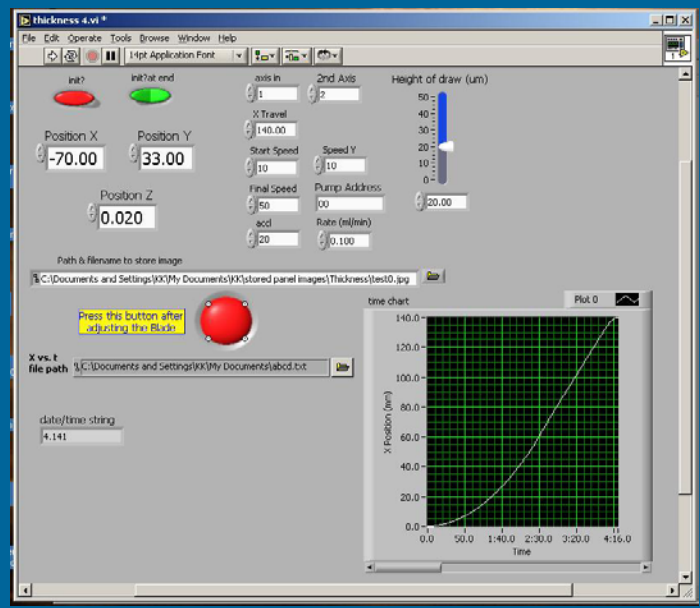
Image object with its conditions shows up on user's desktop almost instantaneously

National Institute of Standards and Technology

Phase 1: Automated Data Entry and Retrieval

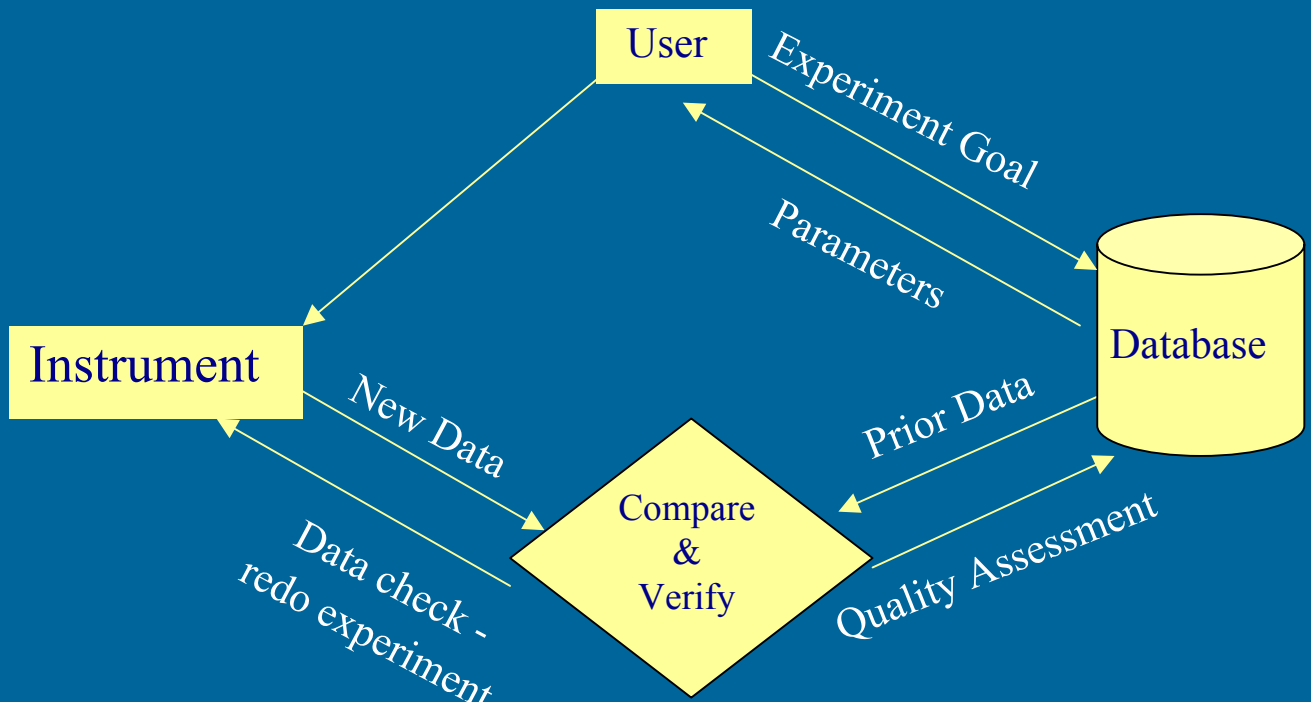




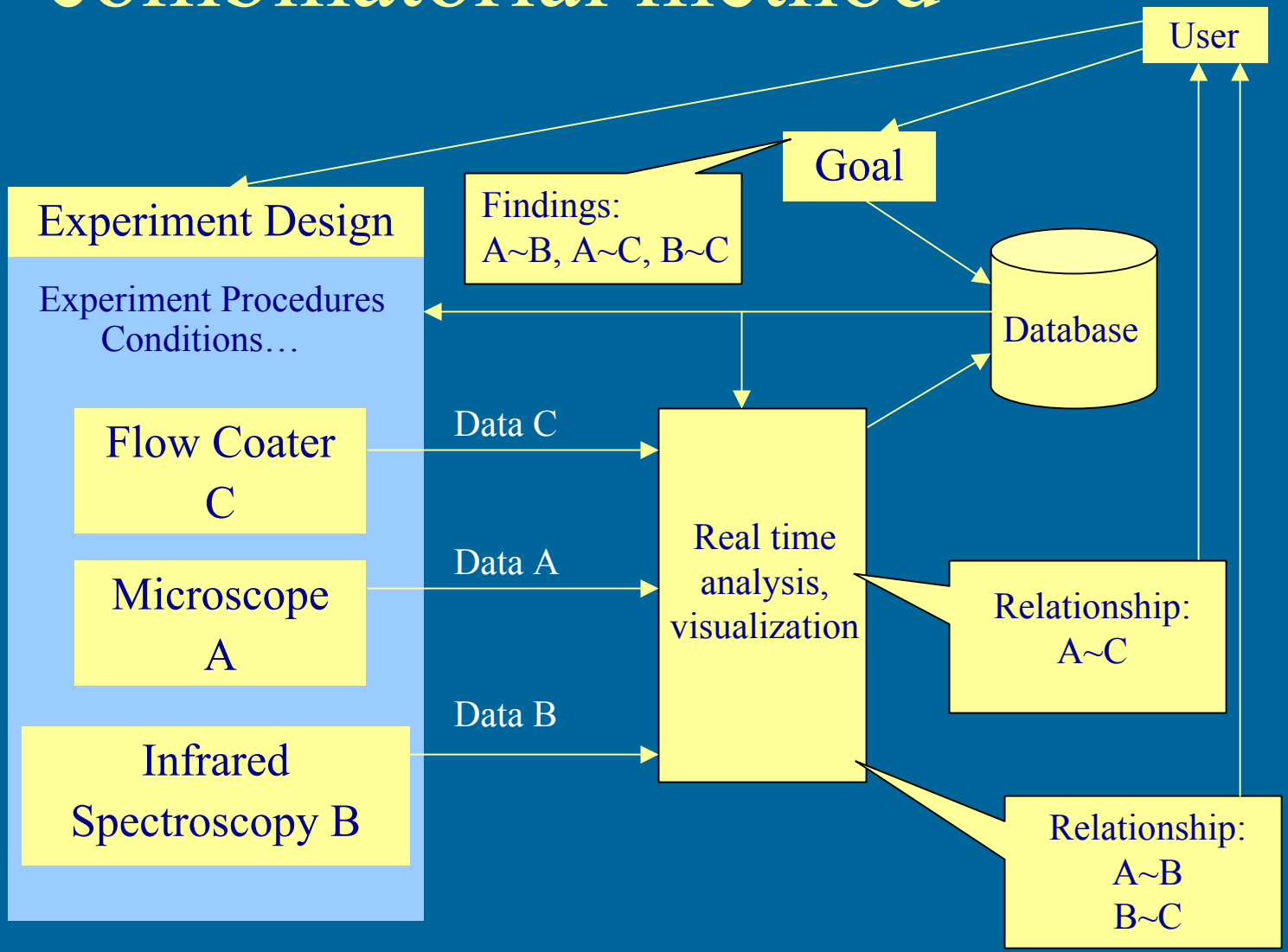


National Institute of Standards and Technology

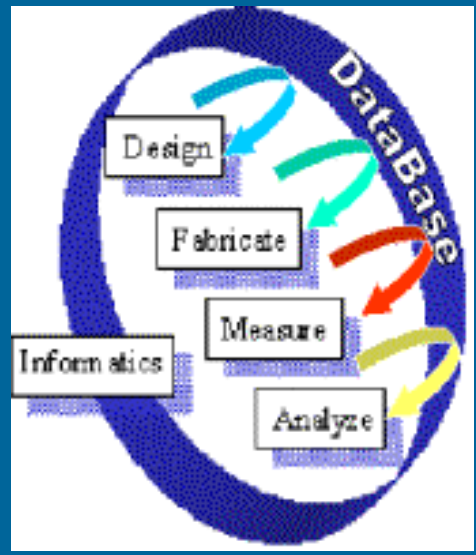
Phase 2: "Expert System" for Each Instrument

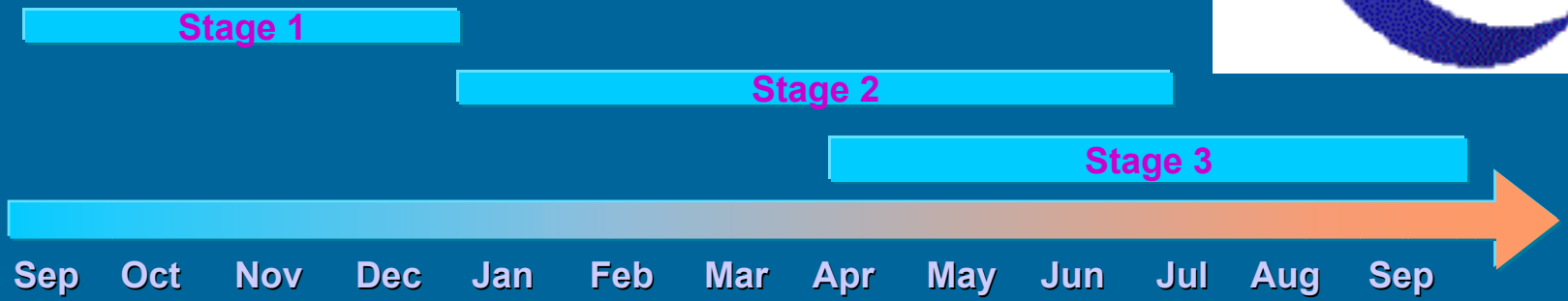


Phase 3: Closing the loop of the combinatorial method



1-Year Schedule





Stage 1: Database setup and programming, e-notebook

Stage 2: Machine programming and automatic data entry and retrieval

Stage 3: Analytic software and feedback automation

NCMC Contributors

- NIST Combinatorial Methods Center

Cher Davis	Alamgir Karim	
Eric Amis	Mike Fasolka	
Kathryn Beers	Christopher Stafford	Polymers Division, NIST
Aaron Forster	Christopher Harrison	
Jeremy Walls	Joao Cabral	

- NIST Combi Informatics Project

NCMC Team	Wenhua Zhang	Polymers Division, NIST
Andrew Roosen	James Warren	Center for Theoretical and Computational Materials Science, NIST
Walter Liggett	James Filliben	Statistical Engineering Division, NIST
John Lu	Nell Sedransk	

National Institute of Standards and Technology

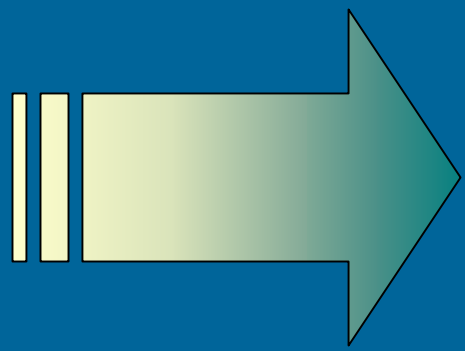
NCMC Combi Informatics Workshop

In conjunction with CTCMS and the Statistical Engineering Division

May 22-23, 2003, NIST Gaithersburg, MD

Informatics Seminar

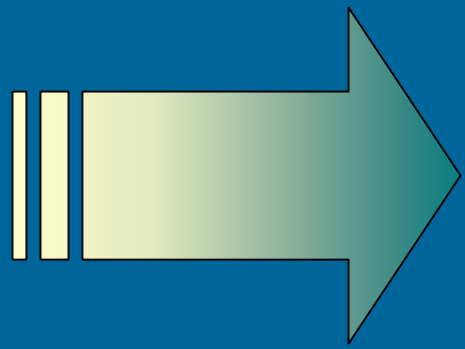
- Resource Integration
- DOE strategies
- Combi Error Analysis



Issues Defined

Informatics Standards Forum

- Research Industry
- Instrument Producers
- Software Providers



Issues Discussed

Visit www.nist.gov/combi for more details

NIST National Institute of Standards and Technology